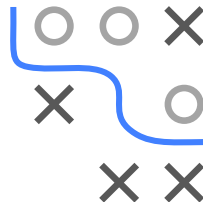
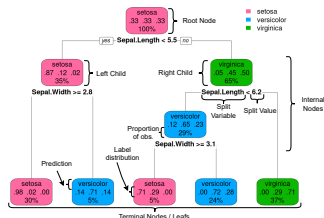


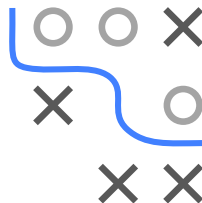
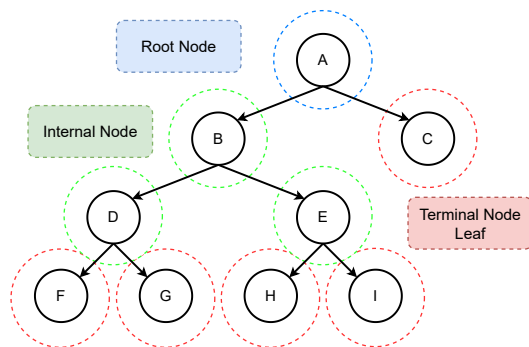
Predictions with CART



- Understand the basic structure of a tree model
- Understand that the basic idea of a tree model is the same for classification and regression
- Understand how the label of a new observation is predicted via CART
- Know hypothesis space of CART

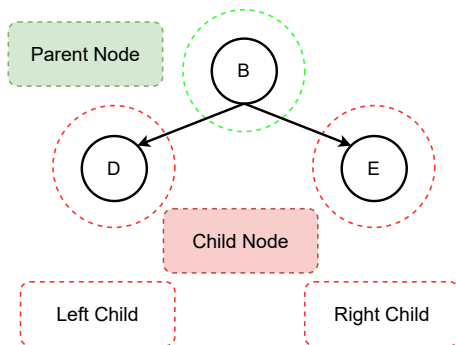


BINARY TREES



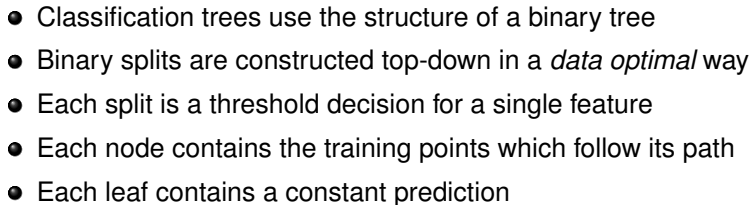
- Binary trees represent a top-down hierarchy with binary splits
- A tree is composed of different node types: a) the root node, b) internal nodes, and c) terminal nodes (also leaves).

BINARY TREES



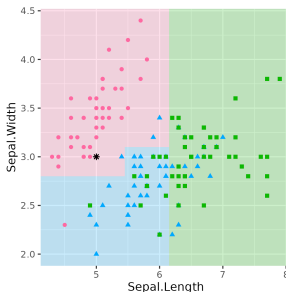
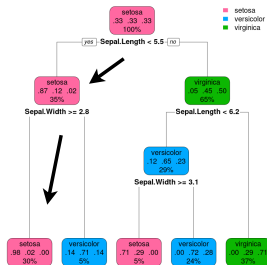
- Nodes have relative relationships, they can be:
 - Parent nodes
 - Child nodes
- Root nodes don't have parents – leaves don't have children

A 3x3 grid with a blue path starting at the top-left corner (0,0) and ending at the bottom-right corner (2,2). The path consists of the cells (0,0), (0,1), (1,1), (1,2), and (2,2). The other cells in the grid contain either a grey circle or a black cross.



CLASSIFICATION TREE MODEL AND PREDICTION

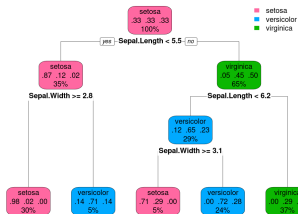
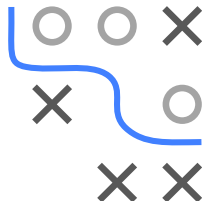
- When predicting new data (here*: `Sepal.Length = 5`, `Sepal.Width = 3`) we use the learned split points and pass an observation through the tree
- Each observation is assigned to exactly one leaf
- Classification trees can make hard-label predictions (here: `setosa`) or predict probabilities (here: 0.98, 0.02, 0.00)



CART AS A RULE BASED MODEL

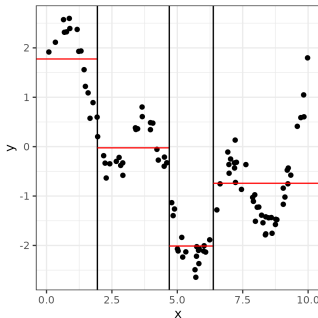
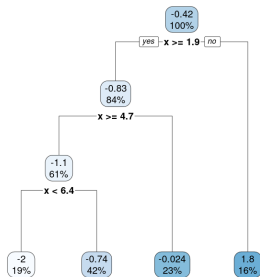
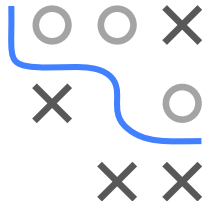
Leaf nodes can be expressed by a set of rules (left to right):

Hard label prediction	Label distribution	Sepal.Width	Sepal.Length
setosa	0.98, 0.02, 0.00	≥ 2.8	< 5.5
versicolor	0.14, 0.71, 0.14	< 2.8	< 5.5
setosa	0.71, 0.29, 0.00	≥ 3.1	≥ 5.5 & < 6.2
versicolor	0.00, 0.72, 0.28	< 3.1	≥ 5.5 & < 6.2
virginica	0.00, 0.29, 0.71	—	≥ 6.2



REGRESSION TREE MODEL AND PREDICTION

- Works the same way as for classification
- But predictions in leaf nodes are a numerical scalar



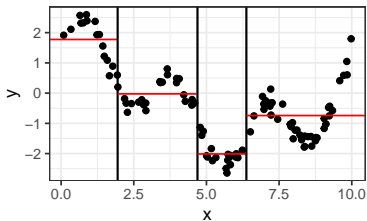
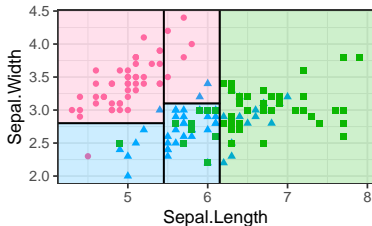
TREE AS AN ADDITIVE MODEL

Trees divide the feature space \mathcal{X} into **rectangular regions**:

$$f(\mathbf{x}) = \sum_{m=1}^M c_m \mathbb{I}(\mathbf{x} \in Q_m),$$

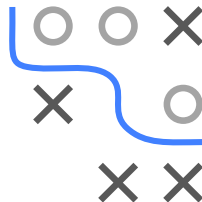
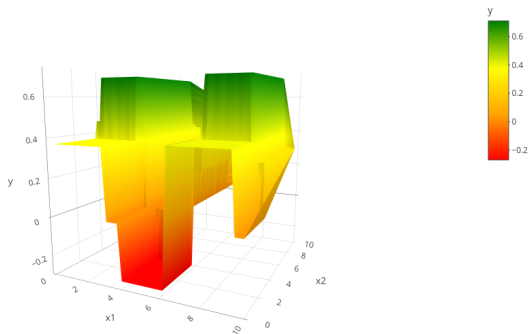
where a tree with M leaf nodes defines M “rectangles” Q_m .

c_m is the predicted numerical response, class label or class distribution in the respective leaf node.



TREE AS AN ADDITIVE MODEL

A 2D regression example:



(For binary classification with probabilities, 2D surface looks similar.)